

Matemáticas Aplicadas a las Ciencias Sociales II: 2º Bachillerato

Capítulo 8: Estimación. Intervalos de confianza

www.apuntesmareaverde.org.es



Autora: Raquel Caro

Revisora: Elena Ramírez

Ilustraciones: Banco de Imágenes de INTEF

Índice

Resumen

Para conocer la opinión de una población sobre el partido político al que piensan votar, se selecciona una muestra adecuadamente, se estudia, y se induce lo que va a votar toda la población. La inferencia estadística, intervalos de confianza y contraste de hipótesis se utilizará para, de los datos que nos suministra una muestra, ser capaces de inducir conclusiones sobre la población. Por ejemplo:

> *Preguntamos a una muestra a qué partido político tiene intención de voto, e inducimos el partido que ganará las elecciones.*

Para hacer control de calidad en un proceso de producción, para ajustar y programar los semáforos en un cruce, para determinar la capacidad curativa de un medicamento… se usa el mismo sistema, se selecciona una muestra. Las conclusiones no pueden ser del tipo: "*Esto va a ser así*" sino que serán probabilísticas: "*Esto va a ser así con tal probabilidad*" o "*Esto va a ser a ser con tal nivel de confianza*".

En los capítulos anteriores has utilizado frecuencias, ahora vamos a asignar probabilidades y al estudiar las distribuciones de probabilidad podremos construir modelos que reflejen la realidad y afirmar, con tal probabilidad, tal nivel de confianza o tal certeza, lo que va a ocurrir.

1. MUESTREO ESTADÍSTICO

Mediante la ***inferencia estadística*** se intenta conocer algo acerca de las características de la población en su conjunto mediante la generalización de lo obtenido en la muestra. Pero es necesario ser consciente de que, en la mayoría de los casos, la verdadera naturaleza y características exactas de la población van a ser desconocidas, y nunca van a poderse conocer con exactitud. A lo más que se puede llegar es a un conocimiento aproximado, que se pretenderá que sea lo más exacto y objetivo posible, dado el nivel de información empírica del que se disponga. Es por ello por lo que la inferencia proporciona conclusiones sin certeza total, sino en términos de probabilidad o de nivel de confianza.

Algunas de las características desconocidas de la población pueden ser su distribución de probabilidad y, en muchos casos, el valor de los parámetros que definen dicha distribución. Así, muchos de los procedimientos básicos de la inferencia estadística clásica están centrados alrededor del valor de dichos parámetros. A continuación desarrollamos la metodología de *estimación de parámetros*.

En muchas ocasiones se desea *estimar* un resultado. Resolver la forma mejor de hacerlo es toda una parte de la Estadística, la *Teoría de Muestras*, que nos indica varios detalles a tener en cuenta:

- ¿Cómo se deben elegir los elementos de la muestra?
- ¿Cuál debe ser el tamaño de la muestra?
- ¿Hasta qué punto la muestra es representativa de la población?

Si se da como resultado de la estimación un valor numérico concreto se habla entonces de ***estimación puntual***, mientras que si se da un conjunto de valores, entre los cuales se espera que se encuentre el verdadero valor del parámetro con un cierto grado de confianza, se habla entonces de ***estimación por intervalo***.

Poniendo otro ejemplo, supongamos que en una estación de ferrocarril se encuentra una máquina automática de café regulada de tal forma que se está interesado en conocer la "cantidad media de café que la máquina suministra en cada taza". Esa "cantidad media de café" es un parámetro poblacional y, por tanto, su valor exacto es desconocido y siempre lo será. Sin embargo, mediante la información muestral, es posible estimar, esto es, ofrecer una aproximación numérica a dicho valor paramétrico desconocido. En este caso, un posible ***estimador puntual*** de la media de la población puede ser la media de la muestra. Si se realiza la ***estimación por intervalo*** se obtiene con una confianza determinada, que la "cantidad media de café" suministrada por taza estará entre dos valores numéricos determinados.

A la hora de estimar parámetros poblacionales, parece una buena estrategia inicial utilizar el que aquí se denominará criterio de analogía. Según este criterio, se elige como estimador de un parámetro poblacional (con significado conocido) su correspondiente análogo en la muestra.

En esta primera sección de este capítulo vamos a *estimar* el valor de un estadístico de una muestra conociendo la población.

En la siguiente haremos algo más útil, estimar el valor de un parámetro de una población, la media o la proporción, a partir del obtenido en una muestra. Conocer el valor exacto va a ser imposible, por eso estudiaremos los intervalos de confianza que nos dirán, con un nivel de confianza un intervalo en el que puede estar el parámetro de la población.

En la tercera sección estudiaremos el contraste de hipótesis.

1.1. Población y muestra

En cursos anteriores ya has estudiado lo que se entiende por muestra y por población:

Definición:

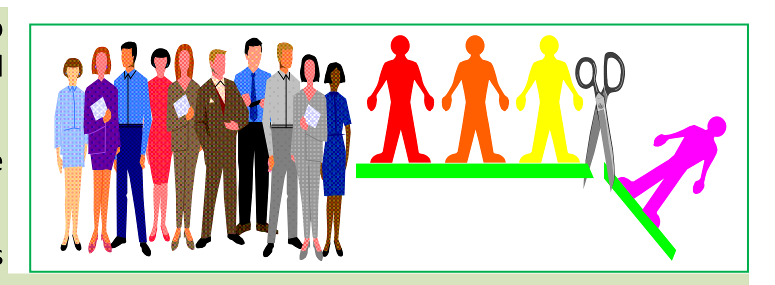
Población estadística, colectivo o universo es el conjunto de todos los individuos (personas, objetos, animales, etc.) que contengan información sobre el fenómeno que se estudia.

Ejemplos:

- Si estudiamos el precio de la vivienda en una ciudad, la población será el total de las viviendas de dicha ciudad.
- Se va a realizar un estudio estadístico sobre el porcentaje de personas casadas en la península. Para ello no es factible estudiar a todos y cada uno de los habitantes por razones de coste y de rapidez en la obtención de la información. Por lo tanto, es necesario acudir a examinar sólo una parte de esta **población**. Esa parte es la **muestra** elegida.

Definición:

Muestra es un subconjunto representativo que se selecciona de la población y sobre el que se va a realizar el análisis estadístico.



Muestreo es el proceso mediante el cual se selecciona la muestra de la población.

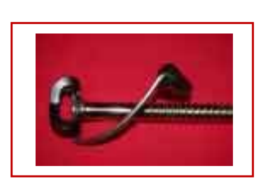
El **tamaño de la muestra** es el número de sus elementos.

Cuando la muestra comprende a todos los elementos de la población, se denomina **censo.**

Ejemplo:

- Si se estudia el precio de la vivienda de una ciudad, lo normal será no recoger información sobre todas las viviendas de la ciudad (ya que sería una labor muy compleja y costosa), sino que se suele seleccionar un subgrupo (muestra) que se entienda que es suficientemente representativo.
- En control de calidad, por ejemplo, si se estudia la vida de un electrodoméstico, y para ello deben funcionar hasta que se estropeen, es absurdo estudiar todos los electrodomésticos (población) pues nos quedamos sin fabricación, por lo que es imprescindible seleccionar una muestra que sea representativa de la población.

Actividades propuestas



1. Señala en qué caso es más conveniente estudiar la población o una muestra:

   a) El diámetro de los tornillos que fabrica una máquina diariamente.
   b) La altura de un grupo de seis amigos.

2. Se puede leer el siguiente titular en el periódico que publica tu instituto: "*La nota media de los alumnos de 2º de Bachillerato de la Comunidad de Madrid es de* 7'9". ¿Cómo se ha llegado a esta conclusión? ¿Se ha estudiado a toda la población? Si hubieran seleccionado para su cálculo solo a las mujeres, ¿sería representativo su valor?

Recuerda que:

La **media muestral** la representamos por $\bar{x}$ o por la letra m, y se define como:

$$\bar{x} = \frac{\sum_i x_i n_i}{n} = \sum_{i=1}^{k} x_i f_i$$

La **desviación típica muestral** la representamos por la letra s, y se define como:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

La media muestral y la desviación típica muestral son los **estadísticos** de la muestra que vamos a usar.

La **media poblacional**, o la media de una distribución, la representamos por la letra griega μ y se define:

$$\mu = E(x) = \sum_i x_i \cdot p(x_i)$$

$$\mu = E(x) = \int_a^b x \cdot f(x) \cdot dx = 0$$

La **desviación típica poblacional**, o de una distribución, la representamos por la letra griega σ y se define:

$$\sigma^2 = \sum_i (x_i - \mu)^2 \cdot p(x_i) = E(x^2) - E^2(x) \qquad \sigma = \sqrt{E(x^2) - E^2(x)}$$

$$\sigma^2 = \int_a^b (x-\mu)^2 \cdot f(x) \cdot dx = 0$$

La media poblacional y la desviación típica poblacional son los **parámetros** de la población que vamos a usar.

Recuerda que:

Estadístico: valor obtenido de la muestra.

Parámetro: valor de la población.



DISTRIBUCIÓN DE PROBABILIDAD. Distribución y tipos de distribuciones. Estadística. Yasuri May

https://www.youtube.com/watch?v=AJt01UoVWZc

1.2. Tipos de muestreos aleatorios

La forma de seleccionar la muestra, **muestreo**, debe reunir unas determinadas características para que pueda caracterizar a la población, ser representativa de la población. Debe ser un muestreo **aleatorio**, es decir, al azar. Todos los individuos de la población deben tener las mismas posibilidades de ser seleccionados para la muestra.

Ejemplos:

- *Se quiere estudiar el nivel adquisitivo de los personas de una ciudad, para lo que pasamos una encuesta a la puerta del Corte Inglés, ¿te parece un muestreo aleatorio?*

No lo es. Las personas que entran en un determinado establecimiento no representan a toda la población.

- *Vas a hacer un estudio sobre los gustos musicales de los jóvenes, y para ello, preguntas a cinco de entre tus amistades, ¿te parece un muestreo aleatorio?*

No lo es. Tus amistades pueden tener unos gustos diferentes a los del resto de la población.

Si la muestra está mal elegida, no es representativa, se producen sesgos, errores en los resultados del estudio.

Hay muchos tipos de muestreo, que darían para analizar en un libro sobre "Muestreo". Pero es conveniente conocer alguno:

Muestreo aleatorio simple

Todos los individuos de la población tienen la misma probabilidad de ser elegidos en la muestra.

Muestreo aleatorio sistemático

Se ordenan los individuos de la población. Se elige al azar un individuo, y se selecciona la muestra tomando individuos mediante saltos igualmente espaciados.

Muestreo aleatorio estratificado

Se divide la población en grupos homogéneos de una determinada característica, *estratos*, por ejemplo edad, y se toma una muestra aleatoria simple en cada estrato.

Ejemplo:

- Se estudia el estado de los huesos de la población de un país, y se divide la población en "niños", "jóvenes", "edad media" y "tercera edad". En cada grupo se hace un muestreo aleatorio simple.

Muestreo por conglomerados o áreas

Se divide la población en conglomerados o áreas, selecciona al azar uno o varios conglomerados y se estudia.

Ejemplo

- Se estudia la incidencia de enfermedades cardiacas en la población rural española. Para ello se hace un censo de pueblos y se eligen varios al azar, donde se estudia a la población

Muestreo no aleatorio

A veces también se usa. ***Por ejemplo,*** conoces la estimación de voto que suele hacerse a pie de urna. Es cómodo, barato pero no es representativo.

1.3. Tamaño y representatividad de una muestra

Cuando se elige una muestra los dos aspectos que hay que tener en cuenta son, el tamaño y la representatividad de la muestra. Por tanto, es importante elegir una muestra adecuada que represente fielmente la población de interés y tenga un tamaño suficiente para obtener resultados fiables

Si la muestra es demasiado pequeña, aunque esté bien elegida, el resultado no será fiable.

Ejemplo:

- *Queremos estudiar la estatura de la población española. Para ello elegimos a una persona al azar y la medimos.*

Evidentemente este resultado no es fiable. La muestra es demasiado pequeña.

Si la muestra es demasiado grande los resultados serán muy fiables, pero el gasto puede ser demasiado elevado. Incluso, en ocasiones, muestras demasiado grandes no nos proporcionan mejores resultados. Vamos a aprender a encontrar cuál es el tamaño adecuado para que podamos afirmar que la población tiene tal característica con una probabilidad dada, grande.

Cuando una muestra tenga el tamaño adecuado, y haya sido elegida de forma aleatoria diremos que es una muestra representativa.

Si la muestra no ha sido elegida de forma aleatoria diremos que la muestra es **sesgada**.

Actividad resuelta

- *Indica si es población o muestra:*

1) En una ganadería se mejora el pienso de todas las ovejas con un determinado tipo de grano.
2) En otra ganadería se seleccionan 100 ovejas para alimentarlas con ese tipo de grano y estudiar su eficacia.

En el primer caso, todas las ovejas, son la población. En el segundo se ha elegido una muestra.

- *En una serie de televisión tienen dudas sobre qué hacer con la protagonista, si que tenga un accidente o si debe casarse. Van a hacer una consulta. ¿A toda la población o seleccionado una muestra representativa?*

Observa que no sabemos bien cuál sería la población, ¿los que ven esa serie? o ¿toda la población española? Si son los que ven la serie, ¿cómo los conocemos? ¿Cómo preguntar a todos? Parece más operativo preguntar a una muestra.

- *El estudio de la vida media de unas bombillas, ¿se puede hacer sobre toda la población?*

El estudio es destructivo. Si se hiciera sobre toda la población nos quedamos sin bombillas. Es imprescindible tomar una muestra.

Actividades propuestas

3. Para estudiar el número de accidentes de una población de mil conductores, de los cuales la mitad tiene carnet de conducir entre 5 y 20 años, la cuarta parte lo tiene más de 20 años y la otra cuarta parte lo tiene menos de 5 años. Se quiere elegir por muestreo aleatorio estratificado proporcional, 50 conductores, ¿cuántos seleccionarías de cada grupo?

1.4. Teorema central del límite

Cuando el curso pasado estudiamos la distribución normal ya comentamos que se pensó que todos los fenómenos se ajustaban a esa distribución, con la broma de que los matemáticos pensaban que los físicos lo habían comprobado experimentalmente, y los físicos que los matemáticos lo habían demostrado.

Este ajuste de los fenómenos a la distribución normal se conoce como Teorema Central del Límite, que fue enunciado por primera vez por el matemático francés que ya conoces por el cálculo de probabilidades, *Pierre Simon Laplace* (1749 – 1827) y demostrado por el matemático ruso *Alesksandr Mikhailovich Lyapunov* (1857 – 1918).

Teorema Central del Límite:

Sean X_n unas variables aleatorias independientes de una población de media μ finita y desviación típica σ finita. Entonces: La distribución de la media muestral de tamaño n tiene de media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$, y se aproxima a una distribución normal a medida que crece el tamaño de la muestra.

El problema es que no especifica qué se entiende por "crecer el tamaño".

Aunque sí sabemos que si la población de partida es normal, entonces la distribución de las medias muestrales es también normal.

Si la población de partida no es normal entonces la distribución de la media muestral se aproximará a una normal cuando el tamaño de la muestra sea suficientemente grande y las variables aleatorias sean independientes. Vamos a considerar que ese tamaño es grande si es mayor que 30.

Actividad resuelta

- *Los parámetros de una distribución son* $\mu = 10$ *y desviación típica* $\sigma = 20$*. Se extrae una muestra de* 100 *individuos. Calcula* $P(8 < \bar{x} < 12)$.

Por el teorema Central del Límite sabemos que la media muestral de una población normal se distribuye según otra distribución normal $N(\mu, \frac{\sigma}{\sqrt{n}}) = N(10, 20/10) = N(10, 2)$.

Para calcular la probabilidad pedida, tipificamos y buscamos en la tabla de la normal.

$$P(8 < \bar{x} < 12) = P(\frac{8-10}{2} < z < \frac{12-10}{2}) = P(-1 < z < 1) = P(z < 1) - P(z < -1) =$$
$$2P(z < 1) - 1 = 2(0'8416) - 1 = 0'6832$$

Debes recordar para hacerlo las propiedades de la curva normal, el uso de la tabla y cómo se calculan probabilidades con ella.

Actividades propuestas

4. Los parámetros de una distribución son $\mu = 20$ y desviación típica $\sigma = 3$. Se extrae una muestra de 400 individuos. Calcula $P(19.9 < \bar{x} < 20.3)$.



1.5. Distribución de la media muestral

De una población se selecciona una muestra y se calcula su media $\bar{x}$ y su desviación típica, s.

Elegimos otras muestras de la misma población, y de cada una obtenemos su media y desviación típica.

¿Cómo es la distribución de esas medias? ¿Y de esas desviaciones típicas?

Las diferentes medias dan lugar a una variable aleatoria que la vamos a representar por $\overline{X}$.

El **Teorema Central del Límite** nos garantiza que, si las variables son independientes:

La media de la variable aleatoria $\overline{X}$ es la media poblacional μ.

La desviación típica de la variable aleatoria $\overline{X}$ es $\frac{\sigma}{\sqrt{n}}$, donde σ es la desviación típica poblacional y n es el tamaño de las muestras elegidas.

Para valores de n suficientemente grandes, ($n \geq 30$) la distribución de $\overline{X}$ se aproxima a una normal:

$$N(\mu, \frac{\sigma}{\sqrt{n}}).$$

Esta afirmación es cierta, sea cual sea la distribución de la población de partida, tanto si es discreta como si es continua, tanto si es normal (entonces se aproxima a esta normal para valores de n menores que 30) como si no lo es.

Actividad resuelta

- ***Control de las medias muestrales****: En el control de calidad de una fábrica de latas de atún, se envasan latas de* 100 *gramos con una desviación típica de* 2 *gramos. Se empaquetan en cajas de* 50 *latas. Calcula la probabilidad de que la media de las latas de una caja sea menor que* 99 *gramos.*

Los datos que nos dan son la media poblacional, μ = 100, la desviación típica poblacional, σ = 2, y el tamaño de la muestra, n = 50.

Sabemos que la media muestral se distribuye según una $N(\mu, \frac{\sigma}{\sqrt{n}}) = N(100, 0.28)$. Vamos a recordar como calculábamos esas probabilidades.

Queremos calcular $P(\bar{x} < 99)$.

Lo primero tipificamos para pasar a una distribución $N(0, 1)$.

$$P(\bar{x} < 99) = P(z < \frac{99-100}{0.28}) = P(z < -3.54) = 1 - P(z < 3.54)$$



Recuerda:

La distribución normal es simétrica, por eso en la tabla no aparecen valores negativos, pues los calculamos usando los positivos. Buscamos en la tabla 3.54 y obtenemos que $P(z < 3.54) = 0.9998$.

$P(\bar{x} < 99) = 1 - P(z < 3.54) = 1 - 0.9998 = 0.0002$, una probabilidad muy pequeña.

Actividad resuelta

- ***Control de la suma:*** *En el mismo ejemplo anterior determina la probabilidad de que un lote de* 400 *latas pese más de* 40100 *gramos.*

Como la media muestral es igual a $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, entonces $\sum_{i=1}^{n} x_i = n\bar{x}$, por lo que su distribución es una normal de media $n\mu$ y desviación típica $n\frac{\sigma}{\sqrt{n}} = \sigma\sqrt{n}$: $N(n\mu, \sigma\sqrt{n})$.

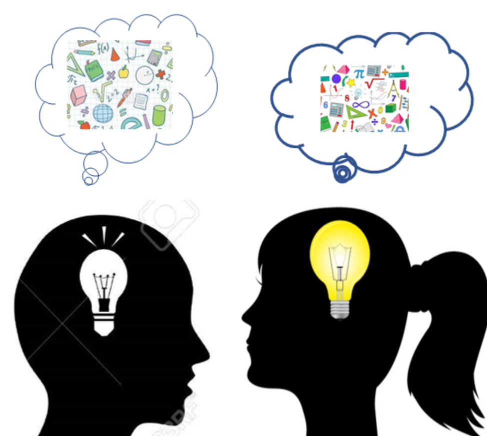
En nuestro caso $N(n\mu, \sigma\sqrt{n}) = N(400\cdot 100, 2\sqrt{400}) = N(40000, 40)$

Queremos calcular

$$P\left(\sum_{i=1}^{n} x_i > 40100\right) = P\left(z > \frac{40100-40000}{40}\right) = P(z > 2.5) = 1 - P(z < 2.5) = 1 - 0.9938 = 0.0062$$

Unas 6 cajas de cada mil pesarán más de 40.1 kg.

Actividades propuestas



5. Los pesos de las ovejas de una cierta ganadería tienen una media de 50 kg con una desviación típica de 4. Elegimos al azar una muestra aleatoria simple de 100 ovejas. A) Determina la probabilidad de que su media sea superior a 51 kg. B) Sea inferior a 56 kg. C) Sea superior a 48 kg. D) Esté entre 48 kg y 52 kg.

6. Una población tiene una media $\mu = 400$ y una desviación típica $\sigma = 20$. Extraemos una muestra de 1000 individuos. Halla el intervalo característico, para una probabilidad de 0.95, de la media muestral. Lo mismo para una probabilidad del 0.99.

7. El peso de una población se estima que tiene de media $\mu = 70$ kg y una desviación típica $\sigma = 10$. Se elige una muestra aleatoria simple de 100 individuos y se pesan todos juntos. Calcula la probabilidad de que dicho peso sea superior a 7010 kg.

1.6. Distribución de una proporción muestral

Hemos estudiado el curso pasado la distribución binomial. Era una situación en que las únicas posibilidades eran "éxito" y "no éxito". Queremos saber cómo se distribuye la proporción muestral (número de éxitos entre el número de veces que se repite el experimento). Cada muestra que obtengamos de tamaño n se distribuye según una distribución binomial $B(1, p)$, por tanto la suma de n variables B(1, p) es una binomial B(n, p) por el principio de reproductividad de la distribución.

Por el Teorema Central del Límite se puede afirmar que la distribución de la proporción muestral $\hat{p}$:

Media: $\mu = p$.

Desviación típica: $\sigma = \sqrt{\frac{p(1-p)}{n}}$

A medida que crece n la distribución de la proporción muestral, $\hat{p} = \frac{x}{n}$, se aproxima a una normal $N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$, siempre que p no tome valores próximos a 0 o a 1.

Actividad resuelta

- *Una envasadora detecta que el 5 % de los paquetes de kilo de arroz tienen exceso de peso. Toman una muestra de 50 paquetes. ¿Qué distribución sigue la proporción de paquetes con exceso de peso? Calcula la probabilidad de que en la muestra elegida existan más de un paquete con exceso de peso.*

La proporción sigue, para n grande, una distribución:

$$N\left(p, \sqrt{\frac{p(1-p)}{n}}\right) = N\left(0.05, \sqrt{\frac{0.05 \cdot 0.95}{50}}\right) = N(0.05,\ 0.03)$$

Como nos piden que haya más de 1 paquete con exceso de peso en la muestra, la proporción de paquetes con exceso de peso es $\frac{1}{50} = 0.02$. Calculamos la probabilidad y tipificamos:

$$\mathrm{P}(\hat{\mathrm{p}} > 0.02) = \mathrm{P}\left(\mathrm{z} > \frac{0.02 - 0.05}{0.03}\right) = \mathrm{P}(\mathrm{z} > -1) = \mathrm{P}(\mathrm{z} < 1) = 0.8413$$

Actividades propuestas

8. En los exámenes de selectividad la proporción de aprobados es del 98 %. Un centro escolar presenta a 78 estudiantes al examen.

a) ¿Qué distribución sigue la proporción de aprobados?

b) Calcula la probabilidad de que en la muestra elegida haya menos de 3 suspensos.

c) Calcula la probabilidad de que en la muestra elegida haya más de 10 suspensos.

d) Calcula la probabilidad de que en la muestra elegida no haya ningún suspenso.

9. En una fábrica de bombillas de bajo consumo hay que rechazar por defectos al 2 % de la producción. Se toma una muestra aleatoria simple de 100 bombillas.

a) ¿Qué distribución sigue la proporción de bombillas defectuosas?

b) Calcula la probabilidad de que en la muestra elegida haya menos de 5 bombillas defectuosas.



2. INTERVALOS DE CONFIANZA

2.1. Estimadores puntuales

En el apartado anterior hemos obtenido información sobre las muestras aleatorias extraídas de una población conocida. Pero es más usual querer obtener información de la población a partir de la información suministrada por una muestra.

Son varios los procesos posibles a seguir: estimación puntual o por intervalos de parámetros, contraste de hipótesis…

Deseamos conocer algo sobre la población, por ejemplo, la media… y para ello se selecciona de forma aleatoria una muestra. En ella podemos calcular esa media… A ese valor lo denominamos **estimador** o **estimador puntual**, y al hecho de hacerlo, una **estimación puntual,** es decir, cuando tenemos la muestra concreta ese estimador tomará un valor concreto.

Con dicha estimación podremos inferir esa media sobre la población. Ya sabemos que no se puede asegurar que la población tenga esa media, sino que la tiene con una cierta probabilidad. Pero al hacerlo así se dice que hemos hecho una estimación puntual.

Todo parámetro poblacional, media, desviación típica, varianza… tiene un estadístico paralelo en la muestra.

Decimos que un estimador es **insesgado** o **centrado** si su media coincide con el valor del parámetro que se quiere estudiar. La media muestral y la proporción muestral son estimadores centrados.

Si no lo es, al error cometido de le denomina **sesgo**.

Un estimador es **eficiente** si su varianza es mínima.

Para medir la eficiencia de un estimador centrado se utiliza la inversa de la varianza.

Ejemplos:

- La media muestral es un estimador centrado de la media poblacional de eficiencia: n/σ^2.
- La proporción muestral es un estimador centrado de la proporción de la población de eficiencia: $\frac{n}{p(1-p)}$.
- Al aumentar el tamaño de la muestra aumenta la eficiencia de la media muestral y de la proporción muestral.

Actividades propuestas

10. Determina la eficiencia de la media muestral si el tamaño de la muestra es 100 y la desviación típica poblacional es 2.

11. Determina la eficiencia de la proporción muestral si el tamaño de la muestra es 100 y la proporción poblacional es 50 %.

2.2. Intervalos de confianza

Ahora queremos, a partir de una muestra de tamaño n, estimar el valor de un parámetro de la población dando un intervalo en el que confiamos que esté dicho parámetro. A este intervalo lo denominamos, **intervalo de confianza**, y se calcula la probabilidad de que eso ocurra a la que se denomina **nivel de confianza**.

Este curso únicamente estudiaremos estimaciones para la media y para la proporción.

Antes de concretarse en un valor para una muestra determinada, cualquier estadístico puede ser tratado como una variable aleatoria cuya distribución de probabilidad dependerá de la distribución de la variable que represente el comportamiento de la población objeto de estudio. Parece razonable aprovechar la distribución de probabilidad del estadístico utilizado como estimador puntual de un parámetro para, basándose en ella, llegar a determinar un intervalo de confianza para el parámetro que se desea estimar. El método que se utiliza para la obtención del intervalo se conoce como ***método del estadístico pivote*** y consta básicamente de los siguientes pasos:

Se elige un estadístico $t(\boldsymbol{X})$, denominado *estadístico pivote*, que cumpla los siguientes requisitos:

- Su expresión debe depender del parámetro θ que se quiere estimar.
- Por último, su distribución de probabilidad ha de ser conocida (y en muchos casos tabulada) y no debe depender del valor de θ.

Para un determinado nivel de confianza, γ, utilizando la distribución de probabilidad de $t(\boldsymbol{X};\theta)$ se calculan los valores $\boldsymbol{k_1}$ y $\boldsymbol{k_2}$, conocidos como **valores críticos**.

En el siguiente apartado se muestran los desarrollos necesarios con vistas a obtener intervalos de confianza para estimar uno de los parámetros de distribución normal, es decir, la media. También se detalla el cálculo de intervalos de confianza para la proporción de éxitos en pruebas binomiales (1, $\boldsymbol{p}$).

Conceptos:

Intervalo de confianza: Si $P(a < X < b) = 0'95$ tenemos el intervalo de confianza (a, b)

Nivel de confianza o coeficiente de confianza: $1 - \alpha = \gamma$, en nuestro ejemplo, 0.95

Nivel de significación o de riesgo: α, en nuestro ejemplo, 0.05

Valor crítico: $\boldsymbol{k_1}$ y $\boldsymbol{k_2}$, que dejan a la derecha (o a la izquierda) un área $\alpha/2$.

En la $N(0, 1)$ son -1.96 y 1.96 para $\alpha = 0.05$.

Margen de error: Diferencia entre los extremos del intervalo de confianza.

Máximo error admisible: Valor prefijado que no puede superar el valor absoluto de la diferencia entre el estimador y el parámetro.

Otros conceptos ya los hemos trabajado:

Población. Parámetro de la población (media, proporción)

Muestra. Estadístico de la muestra. Tamaño de la muestra.

Actividad resuelta

- *Sabemos que en una distribución normal estándar $P(-1.96 < z < 1.96) = 0.95$. Determina un intervalo de confianza con un nivel de confianza del 0.95 de una $N(2, 0.1)$. Determina el margen de error.*

$$P(-1.96 < Z < 1.96) = 0.95 \Rightarrow P(-1.96 < \frac{X-2}{0.1} < 1.96) = 1-\alpha = \gamma = 0.95 \Rightarrow$$

$$P((0.1 \cdot (-1.96)) + 2 < X < (0.1 \cdot 1.96) + 2) = 0.95 \Rightarrow P(1.8 < X < 2.2) = 0.95$$

La variable aleatoria X estará en el intervalo (1.8, 2.2) con un nivel o coeficiente de confianza de 0.95.

El margen de error viene dado por la amplitud del intervalo:

Margen de error: 2.2 – 1.8 = 0.4.



Distribución Normal BACHILLERATO Selectividad Tipificar. Haremos el cuarto ejercicio de selectividad de junio de 2010. En este caso, resolveremos un ejercicio de distribución normal



https://www.youtube.com/watch?v=_08yRdSDbRY

Actividades propuestas

12. Determina un intervalo de confianza con un nivel de confianza del 95 % de una $N(5, 0.01)$. Determina el margen de error.

13. Determina un intervalo de confianza con un nivel de confianza del 99 % de una $N(100, 4)$. Determina el margen de error.

2.3. Intervalo de confianza para la media poblacional con desviación típica conocida

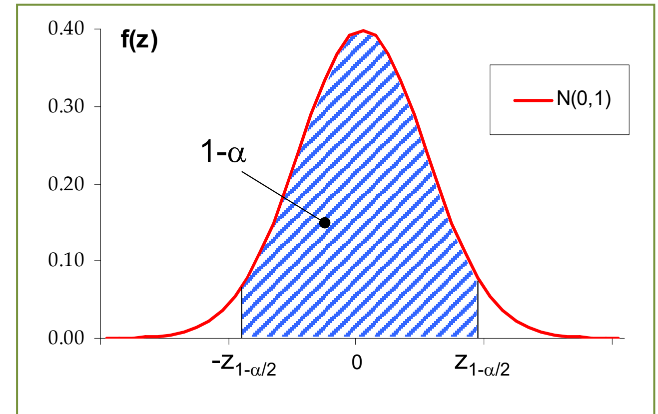
Cuando se quiere construir un intervalo de confianza para estimar la media μ de una población normal en la que se supone que la desviación típica de la distribución, σ, **es conocida**, se utiliza como estimador la media muestral, es decir, se recurre a una muestra de tamaño n de la que se obtiene la media muestral.

Ya sabemos que la media muestral, $\bar{x}$ sigue una distribución normal de media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$ si la población de partida es normal, o si, aunque no lo sea, el tamaño de la muestra es suficientemente grande, $n \geq 30$.

Para obtener, entonces un intervalo de confianza con un nivel de confianza $1 - \alpha = \gamma$ debemos buscar dos valores tales que dividan el área bajo la curva normal en tres zonas, de áreas, $\alpha/2$, $1 - \alpha$ y $\alpha/2$.

$$\phi(-z_{1-\alpha/2}) = \frac{\alpha}{2}; \qquad \phi(z_{1-\alpha/2}) = 1 - \frac{\alpha}{2}$$

La siguiente figura ilustra la localización de estos valores $z_{1-\alpha/2}$ y $-z_{1-\alpha/2}$.



Sabemos que

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}; \qquad \bar{X}: \quad N(\mu, \frac{\sigma}{\sqrt{n}}); \qquad Z: \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}: \quad N(0,\ 1)$$

Se observa que el estadístico depende del parámetro μ que se va a estimar y que su distribución de probabilidad (normal tipificada) es conocida y no depende de dicho parámetro.

Así pues, dado un nivel de confianza 1 – α = γ se buscan dos valores $z_{1-\alpha/2}$ y $-z_{1-\alpha/2}$ que verifiquen:

$$P(-z_{1-\frac{\alpha}{2}} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}) < z_{1-\frac{\alpha}{2}}) = 1-\alpha$$

Llamamos $z_{1-\alpha/2}$ al valor de la $N(0,\ 1)$ que deja un área a la derecha de valor α/2. Entonces, por la simetría de la distribución normal, a la izquierda de $-z_{1-\alpha/2}$ quedará un área igual a α/2. Por tanto:

$$P(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}) = 1-\alpha = \gamma$$

(***Recuerda:*** Si (1 – α)·100 % = 95 %, entonces $z_{1-\alpha/2}$ = 1.96).

A continuación se puede despejar la media poblacional para obtener el intervalo de confianza:

$$P(-z_{1-\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}) = 1-\alpha \Rightarrow P(-z_{1-\frac{\alpha}{2}} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{1-\frac{\alpha}{2}}) = 1-\alpha \Rightarrow$$

$$P(-z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \bar{x}-\mu < z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}) = 1-\alpha \Rightarrow P(|\bar{x}-\mu| < z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}) = 1-\alpha$$

Una vez obtenida la media muestral determinamos, con un nivel de confianza 1 – α = γ el intervalo de confianza. La media poblacional μ, puede pertenecer o no a dicho intervalo.

Por tanto, se obtiene para la media poblacional el intervalo al (1 – α)·100 % de confianza:

$$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

Por último, es interesante recordar que el intervalo de confianza se interpreta de la siguiente manera: si tuviésemos un número infinito de muestras de la población, y construyésemos con cada una un intervalo, entonces el 100·γ % de dichos intervalos contendría al verdadero valor del parámetro μ. En la práctica, sólo tenemos una muestra, y por eso sólo podemos construir un intervalo. No tiene entonces sentido interpretar el intervalo como la región en la que estará μ con probabilidad γ, puesto que en el intervalo calculado, la media μ estará o no estará. Por eso, para expresar nuestra incertidumbre sobre si el intervalo calculado con nuestra muestra contiene o no al parámetro μ emplearemos la expresión *nivel de confianza*.

Actividad resuelta

- *Si se puede realizar la hipótesis de que el consumo de combustible sigue una distribución normal, veamos el intervalo de confianza para la media al* 95 %*, suponiendo conocida la varianza (igual a 7684.3 l^2). Se recoge una muestra aleatoria simple de tamaño 20, y se obtiene una media muestral de* 3937.9 *l*.

Para un nivel de confianza del 95 % la tabla de la normal estándar nos dan que $z_{1-\alpha/2}$ = 1.96.

$$\left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(3937.9 - 1.96 \cdot \frac{87.66}{\sqrt{20}},\ 3937.9 + 1.96 \cdot \frac{87.66}{\sqrt{20}}\right) = (3899.5,\ 3976.3).$$

Actividad resuelta

- *El tiempo de renovación de un teléfono móvil, expresado en años, se puede aproximar mediante una distribución normal con desviación típica* 0.4 *años. Se toma una muestra aleatoria simple de* 100 *usuarios y se obtiene una media muestral igual a* 1.5 *años. Determínese un intervalo de confianza al* 95 % *para el tiempo medio de renovación de un teléfono móvil*.

Buscamos en la tabla de la normal estándar y se obtiene que $z_{1-\alpha/2}$ = 1.96 para un nivel de confianza del 95 %. Conocemos la desviación típica poblacional σ = 0.4, y la muestra nos da una media $\bar{x} = 1.5$.

El intervalo de confianza pedido es:

$$\left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(1'5 - 1'96 \cdot \frac{0'4}{\sqrt{100}}, 1'5 + 1'96 \cdot \frac{0'4}{\sqrt{100}}\right) = (1'5 - 0'0784, 1'5 + 0'0784) = (1'4216, 1'5784)$$

Tenemos la confianza de que el 95 % de los casos la media poblacional pertenecerá al intervalo:

(1.4216, 1.5784).

Actividades propuestas

14. Determina un intervalo de confianza para la media poblacional con un nivel de confianza del 95 % de una población de desviación típica conocida, σ = 2, si hemos escogido una muestra aleatoria simple de tamaño 400 y calculado la media muestral que es 50.5.

15. Determina un intervalo de confianza para la media poblacional con un nivel de confianza del 98 % de una población de desviación típica conocida, σ = 2, si hemos escogido una muestra aleatoria simple de tamaño 400 y calculado la media muestral que es 50.5. Compara con el anterior intervalo de confianza.

16. Se ha tomado una muestra aleatoria simple de 16 pacientes y se ha anotado el número de días que han recibido tratamiento para los trastornos del sueño que sufren. Los resultados han sido:

280; 285; 295; 330; 290; 350; 360; 320; 295; 310; 300; 305; 295; 280; 315; 305.

Se sabe que la duración, en días, del tratamiento se puede aproximar por una variable aleatoria con distribución normal de media μ desconocida y desviación típica 34.5 días. Determina un intervalo de confianza con un nivel del 95 % para la media poblacional.

2.4. Relación entre nivel de confianza, error admisible y tamaño de la muestra

Hemos visto que $P(|\bar{x}-\mu| < z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}) = 1-\alpha$, es decir, el (1 – α)·100 % de las muestras cumplen que:

$$|\bar{x}-\mu| < z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Definición:

Se llama error máximo admisible al valor $E = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$.

Observa que depende del tamaño de la muestra y del nivel de confianza. Al aumentar el tamaño de la muestra disminuye el error máximo admisible, y al aumentar el nivel de confianza también aumenta el error máximo admisible. Puedes comprobarlo con la tabla de la normal estándar, y los niveles de confianza más usados:

1 – α	α	$z_{1-\frac{\alpha}{2}}$
0.90	0.10	1.645
0.95	0.05	1.96
0.99	0.01	2.575

Si nos fijan el error máximo admisible, E, y el nivel de confianza 1 – α, podemos determinar el mínimo tamaño que debe tener la muestra simplemente despejando:

$$E = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow \sqrt{n} = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{E} \Rightarrow n = \left(z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{E} \right)^2.$$

Observa que el tamaño de la muestra debe ser más grande cuanto menor sea el error máximo admisible:

✓ Para estimaciones más precisas se debe aumentar el tamaño de la muestra.

Al aumentar el nivel de confianza 1 – α aumenta el tamaño de la muestra, luego:

✓ Para aumentar el nivel de confianza se debe aumentar el tamaño de la muestra.

Actividad resuelta

- *¿Cuál es el número mínimo de estudiantes que debemos elegir de una población de* σ = 2, *para una muestra aleatoria simple si el error mínimo admisible es de* 0.1, *y el nivel de confianza del* 95 %*?*

$$n \geq \left(z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{E} \right)^2 \Rightarrow n \geq \left(1.96 \cdot \frac{2}{0.1}\right)^2 = 1\,536.64$$

La muestra debe tener al menos 1 537 estudiantes.

Conocido el tamaño de la muestra y el error máximo admisible, despejando y buscando en la tabla, también podemos determinar el nivel de confianza.

$$E = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow z_{1-\frac{\alpha}{2}} = E \cdot \frac{\sqrt{n}}{\sigma}$$

Actividad resuelta

- *El otorrino conoce que la desviación típica del tiempo de respuesta a un sonido es de un segundo. Desea estudiar dicho tiempo de respuesta con un error máximo admisible de* 0.1 *haciendo un estudio con* 100 *pacientes: Determina con qué nivel de confianza obtendrá el intervalo de confianza.*

$$z_{1-\frac{\alpha}{2}} = E \cdot \frac{\sqrt{n}}{\sigma} = 0.1 \cdot \frac{\sqrt{100}}{1} = 1$$

Buscamos en la tabla:

$$P(Z < z_{1-\frac{\alpha}{2}}) = 1-\alpha \Rightarrow P(Z < 1) = 0.8413$$

es decir que el nivel de confianza es del 84.13 %.

Actividad resuelta

- *En la población de estudiantes de desviación típica σ = 2, se quiere pasar una prueba a 100 estudiantes para determinar sus conocimientos de Matemáticas con un error mínimo del 0.5. ¿Cuál es el nivel de confianza obtenido?*

$$z_{1-\frac{\alpha}{2}} = E \cdot \frac{\sqrt{n}}{\sigma} = 0.5 \cdot \frac{\sqrt{100}}{2} = 2.5$$

Buscamos en la tabla:

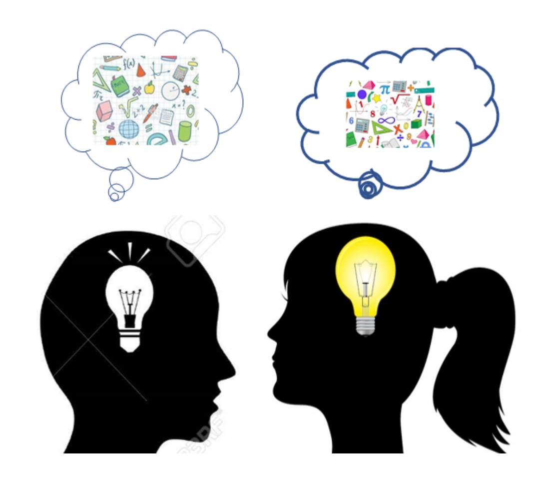
$$P(Z < z_{1-\frac{\alpha}{2}}) = 1-\alpha \Rightarrow P(Z < 2.5) = 0.9938$$

es decir que el nivel de confianza es del 99.38 %.

Actividades propuestas

17. ¿Qué tamaño mínimo debe tener una muestra para que el error máximo cometido en la estimación de la media sea menor de 0,1 unidades, con un nivel de confianza del 95 %, sabiendo que la desviación típica poblacional es conocida y vale 4?

18. Determina el tamaño muestral mínimo necesario para que el valor absoluto de la diferencia entre la media muestral y la media poblacional sea menor o igual a 0.02 años con un nivel de confianza del 90 % sabiendo que la población se distribuye según una normal de desviación típica 0.4.



19. En el estudio anterior se toma una muestra de 49 individuos. Queremos que el error máximo admisible sea de 0.02. ¿Cuál será el nivel de confianza?

El intervalo sobre el valor del parámetro, que se construirá utilizando las propiedades del estimador, se denomina ***intervalo de confianza***.

- ✓ Cuanto más estrecho sea dicho intervalo, menos incertidumbre existirá sobre el verdadero valor del parámetro.
- ✓ Además del concepto de confianza, que se acaba de analizar, en los intervalos aparecen los conceptos de **precisión** y de **amplitud**.

La ***amplitud*** es, la diferencia entre los extremos del intervalo, es decir, $t_S(X) - t_I(X)$.

Para una muestra concreta, la amplitud del intervalo construido a partir de ella será: $t_S(X^0) - t_I(X^0)$.

La ***precisión*** es una forma de evaluar el grado de eficacia del intervalo, y está inversamente relacionado con el concepto de amplitud. En principio será deseable que los intervalos construidos tengan la máxima precisión posible, aunque el tamaño muestral siempre será una limitación, ya que si es muy pequeño, no se puede conseguir una precisión elevada.

Ya se ha dicho que entre precisión y amplitud existe una relación inversa: a mayor precisión deseada, menor ha de ser la amplitud del intervalo construido. Por ello, en principio lo deseable es que el intervalo presente la menor amplitud posible.

Si se obtiene un intervalo a partir de una muestra de tamaño 100, ¿cómo puede mejorarse este intervalo?

- ✓ Una posibilidad es aumentar la precisión. Pero para aumentar la precisión (lo que equivale a disminuir la amplitud), manteniendo el tamaño muestral el único instrumento que existe es el nivel de confianza. Así, es necesario disminuir la confianza (ya que la precisión ha mejorado). Es decir, si la confianza pasa del 99 % a ser, por ejemplo, del 95 %, se puede obtener una amplitud menor.
- ✓ Otra posibilidad es aumentar la confianza. En tal caso, de manera análoga, debería disminuirse la precisión (lo que equivale a aumentar la amplitud).

Si existe la posibilidad de aumentar el tamaño muestral (es decir, si se puede disponer de más información, lo que supone una situación mejor), se puede aumentar la precisión sin modificar la confianza o aumentar la confianza sin modificar la precisión.

Por ejemplo, si se aumenta el tamaño muestral a 200, se puede aumentar la precisión o aumentar la confianza del intervalo, sin modificar la otra característica. Realmente, aumentando el tamaño muestral siempre mejorará el intervalo construido, pero dicho aumento suele tener un coste. Por lo tanto, cuando se quiere construir un intervalo de confianza para un parámetro, antes de obtener la muestra, puede ser interesante realizar un estudio previo para obtener el valor de n óptimo en términos de relación coste-beneficio.

2.5. Intervalo de confianza para la proporción en muestras grandes

La construcción de un intervalo de confianza para la proporción de éxitos en una prueba de Bernoulli se puede llevar a cabo utilizando el estimador puntual que se ha visto en el apartado de estimación puntual. Entonces, se había demostrado que el estimador de la proporción poblacional es un estimador insesgado $\hat{p}$.

Sabemos por el teorema central del límite que la proporción muestral se distribuye según una distribución normal $N(p, \sqrt{\frac{p(1-p)}{n}})$ para n suficientemente grande.

Tipificando la variable obtenemos una distribución $N(0, 1)$, por lo tanto:

Dado un nivel de confianza $1 - \alpha = \gamma$, se pueden buscar dos valores $z_{1-\alpha/2}$ y $-z_{1-\alpha/2}$ que verifiquen:

$$\phi(-z_{1-\alpha/2}) = \frac{\alpha}{2}; \qquad \phi(z_{1-\alpha/2}) = 1 - \frac{\alpha}{2}$$

De manera que construimos el intervalo de confianza para la proporción de éxitos p. La varianza es desconocida y por tanto se utiliza como desviación típica su estimador puntual, $\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}$:

$$P\left(-z_{1-\alpha/2} < \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}} < z_{1-\alpha/2}\right) = 1-\alpha \Rightarrow$$

$$P\left(-\hat{p} - z_{1-\alpha/2}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} < -p < -\hat{p} + z_{1-\alpha/2}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\right) = 1-\alpha \Rightarrow$$

Multiplicamos por -1, y cambiamos el sentido de la desigualdad:

$$P\left(\hat{p} - z_{1-\alpha/2}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}} < p < \hat{p} + z_{1-\alpha/2}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\right) = 1-\alpha \Rightarrow$$

Tenemos el intervalo para la proporción poblacional:

$$p \in \left(\hat{p} - z_{1-\alpha/2}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}},\quad \hat{p} + z_{1-\alpha/2}\cdot\sqrt{\frac{\hat{p}\cdot(1-\hat{p})}{n}}\right)_{\gamma}$$

Con lo que se obtiene el intervalo de confianza para la proporción de éxitos al nivel de confianza, γ. Se puede demostrar que es el intervalo de menor amplitud dado un nivel de confianza.

Actividad resuelta

- *Determina el intervalo de confianza al* 99 % *para la proporción de componentes defectuosos que se producen en una fábrica. Para ello se ha elegido una muestra aleatoria simple de* 1000 *componentes y en ella se ha obtenido que la proporción de defectuosos es del* 3.7 %.



Buscamos en la tabla de la normal el valor de z para una probabilidad de 0.99, y se obtiene 2.58, es decir $z_{1-\alpha/2}$ = 2.58. Conocemos $\hat{p}$ = 0.037, n = 1000, por lo que el intervalo de confianza pedido es:

$$\left(\hat{p} - z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)_{99\%} =$$

$$\left(0.037 - 2.58\sqrt{\frac{0.037 \cdot 0.963}{1000}}, \quad 0.037 + 2.58\sqrt{\frac{0.037 \cdot 0.963}{1000}}\right)_{99\%} = (0.0216, \quad 0.0524)_{99\%}$$

Con un nivel de confianza del 99 % la proporción de defectuosos poblacional está entre 2.16 % y 5.24 %.

Actividad resuelta

- *Determina el intervalo de confianza al nivel de confianza del* 90 % *y del* 99 % *para estimar la proporción enfermos de la gripe en la población si de una muestra de 120 personas hay 20 con gripe. Determina en cada caso el margen de error.*

Buscamos en la tabla de la normal los valores de $z_{1-\alpha/2}$ para esos niveles de confianza y obtenemos para 90 %, $z_{1-\alpha/2}$ = 1.645, y para el 99 %, $z_{1-\alpha/2}$ = 2.575. Conocemos n = 120, $\hat{p}$ = 20/120 = 1/6. Calculamos:

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{\frac{1}{6} \cdot \frac{5}{6}}{120}} = 0.034$$

Por tanto, los intervalos de confianza pedidos son:

$$90\ \% \Rightarrow \left(\hat{p} - z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)_{90\%} =$$

$$\left(\frac{1}{6} - 1.645 \cdot 0.034, \quad \frac{1}{6} + 1.645 \cdot 0.034\right)_{90\%} = (0.111, \quad 0.223)_{909\%}$$

Margen de error = 0.223 – 0.111 = 0.112.

Podemos interpretarlo como que habrá entre un 11 % y un 22 % de personas con gripe.

$$99\ \% \Rightarrow \left(\hat{p} - z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)_{99\%} =$$

$$\left(\frac{1}{6} - 2.58 \cdot 0.034, \quad \frac{1}{6} + 2.58 \cdot 0.034\right)_{99\%} = (0.079, \quad 0.254)_{99\%}$$

Margen de error = 0.254 – 0.079 = 0.175.

Podemos interpretarlo como que habrá aproximadamente entre un 8 % y un 25 % de personas con gripe.

Observa que:

Al aumentar el nivel de confianza, aumenta la amplitud del intervalo y por lo tanto aumenta el margen de error.

Actividades propuestas

20. Determina el intervalo de confianza para la proporción de árboles enfermos en Madrid con un nivel de confianza del 95 %, si se ha elegido una muestra aleatoria simple de 100 árboles de los que hay 20 enfermos.

21. Se quiere estudiar la proporción de estudiantes que hacen actividades extraescolares. Para ello se ha seleccionado una muestra de 400 estudiantes de los cuales 100 hacen actividades extraescolares. Determina el intervalo de confianza para la proporción con un nivel de confianza del 95 %.

2.6. Determinación del tamaño de la muestra para una proporción

Para determinar el tamaño partimos de dos situaciones diferentes

1. Que se conozca la media o la proporción poblaciones
2. Que no se conozca

Ya hemos determinado el tamaño de la muestra para la media poblacional, ahora veremos algún ejemplo para la proporción.

El procedimiento es el mismo que antes. La diferencia va a estar en despejar el tamaño pues vamos a tener una desigualdad con raíces cuadradas. Como el tamaño buscado también es una desigualdad podremos simplificar esa desigualdad.

Veámoslo con unos ejemplos:

Actividad resuelta

- *¿Cuál debe ser el tamaño de la muestra en una población de* 8 *millones de votantes para conocer si tienen la intención de votar a un determinado partido político con una probabilidad de acierto del* 0.95 *y un margen de error inferior a* 0.02*? Se conoce la proporción poblacional: 35 %.*

Utilizamos intervalos de confianza:

Es una distribución binomial, pues un votante o vota a dicho partido, o no lo vota.

Llamamos n al tamaño de la muestra, p al número de los que votarán al partido en la población, X a los que votan al partido en la muestra.

$$P(-0.02 \leq \frac{X}{n} - p \leq 0.02) = P((-0.02 + p) \cdot n \leq X \leq (0.02 + p) \cdot n) \geq 0.95$$

En la distribución binomial tenemos que la media es np y la varianza $npq = np(1-p)$. Pasamos de la distribución binomial a la distribución normal, añadiendo 0.5 de la longitud de los intervalos:

$$P(-0.02 \leq \frac{X}{n} - p \leq 0.02) = P(-0.02n + pn - 0.5 \leq X \leq 0.02n + pn + 0.5) \geq 0.95$$

Tipificamos:

$$P(\frac{-0.02n-0.5}{\sqrt{np(1-p)}} \leq \frac{X-np}{\sqrt{np(1-p)}} \leq \frac{0.02n+0.5}{\sqrt{np(1-p)}}) \geq 0.95 \Rightarrow$$

$$2P(z \leq \frac{0.02n+0.5}{\sqrt{np(1-p)}}) - 1 \geq 0.95 \Rightarrow$$

$$P(z \leq \frac{0.02n + 0.5}{\sqrt{np(1-p)}}) \geq 0.975$$

Buscamos en la tabla de la normal estándar y obtenemos que

$\frac{0.02n+0.5}{\sqrt{np(1-p)}} \geq 1.96.$ (1)

La proporción es conocida p = 0.35, q = 0.65,

$$0.02n + 0.5 \geq 1.96\sqrt{np(1-p)} \Rightarrow 0.02n + 0.5 \geq 1.96\sqrt{n \cdot 0.35 \cdot 0.65}$$

Podemos resolver la desigualdad pero también podemos simplificarla, pues se seguirá verificando para este caso (aunque no en el otro sentido):

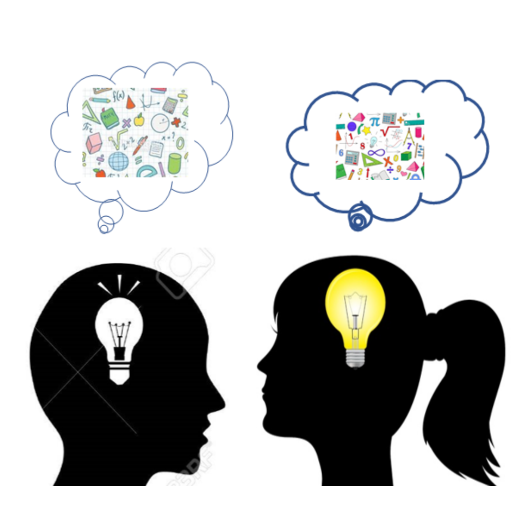
$$0.02n \geq 1.96\sqrt{n \cdot 0.35 \cdot 0.65}$$

Elevamos al cuadrado y despejamos:

$$n \geq 2\,184.91 \Rightarrow n \geq 2\,185.$$

Por tanto se debe pasar la encuesta a 2 185 votantes o más.

Actividad resuelta



- *¿Cuál debe ser el tamaño de la muestra en una población de 8 millones de votantes para conocer si tienen la intención de votar a un determinado partido político con una probabilidad de acierto del* 0.95 *y un margen de error inferior a* 0.02*? Se desconoce la proporción poblacional.*



Es el mismo problema anterior, pero desconocemos la proporción.

Partimos de la desigualdad (1):

$$\frac{0.02n+0.5}{\sqrt{np(1-p)}} \geq 1.96 \Rightarrow 0.02n + 0.5 \geq 1.96\sqrt{np(1-p)}$$

Donde tenemos dos variables n y p. Vamos a acotar $p(1-p)$. Dibujamos la parábola $y = x(1-x)$ que alcanza su valor máximo, 1/4, para x = 1/2, por lo que $p(1-p) \leq 1/4$. Sustituimos este valor.

$$0.02n + 0.5 \geq 1.96\sqrt{np(1-p)} \geq 1.96\sqrt{\frac{n}{4}}$$

Eliminamos 0.5 (para simplificar cálculos), elevamos al cuadrado, y obtenemos que: $n \geq 2\,401$.

La encuesta debe de realizarse para más de 2 401 votantes.

Hemos calculado el tamaño de la muestra con un margen de error no superior a 0.02 y una certeza del 95 %.

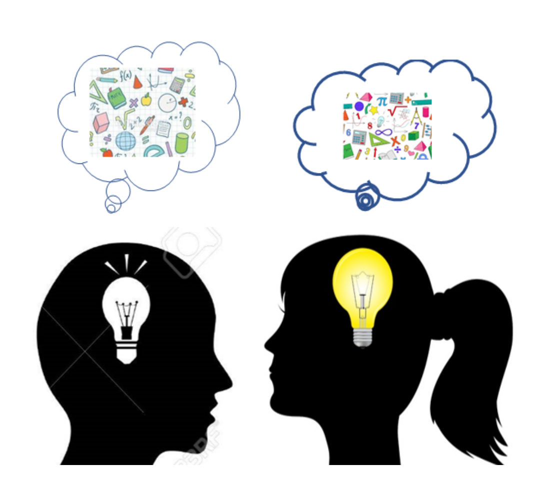
Actividades propuestas

22. ¿Cuántas veces se debe lanzar una moneda para que la proporción de caras no se aparte de la teórica, 1/2, más de una centésima, con un grado de certeza no inferior al 95 %? ¿Cuántas, con el mismo margen de error y una certeza no inferior al 99 %? ¿Lo mismo con 99.9 % de certeza? (*Soluciones*: $n \geq 9\,504$, $n \geq 16\,412$, $n \geq 26\,632$)

Volvemos al problema de las encuestas de votos.

Actividad resuelta

- *En una población de 8 millones de votantes elegimos una muestra aleatoria de 2 000 de la que 700 personas nos afirman que van a votar a un determinado partido. ¿Qué podemos asegurar sobre el número de votos que recibirá dicho partido?*



Como 700/2000 = 35, una primera respuesta podría ser que 0.35·8000000 = 2 800 000 votos, pero ¿qué confianza podemos tener de ese resultado.

Fijamos un nivel de significación α, o un grado de confianza, $1 - \alpha = \gamma$. Sea $\alpha = 0.05$ y $\gamma = 1 - \alpha = 0.95$.

Sea p la proporción de votantes al partido estudiado. Tenemos una distribución binomial de media $\mu = np = 2000 \cdot p$ y $\sigma = \sqrt{npq} = \sqrt{2000 \cdot p(1-p)}$. Calculamos la probabilidad de que el número de votantes al partido estudiado de la muestra sea: $P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 0.95$

Pasamos de la distribución binomial a la normal para calcular k y p:

$$P(\mu - k\sigma - 0.5 \leq X \leq \mu + k\sigma + 0.5) \geq 0.95$$

Tipificamos: $P(\frac{-k\sigma-0.5}{\sigma} \leq Z \leq \frac{+k\sigma+0.5}{\sigma}) \geq 0.95$

Obtenemos que $z = \frac{+k\sigma+0.5}{\sigma} \geq 1.96$, por lo que $k\sigma + 0.5 \geq 1.96\sigma$. Debemos sustituir μ y α en función de p como se hizo anteriormente y se obtiene que: $0.3280 \leq p \leq 0.3719$, es decir que la proporción de votantes debe estar entre el 33 % y el 37 %.

Actividades propuestas

23. Rehaz los cálculos de la actividad anterior para un nivel de confianza del 99 %

24. Se investigan los hábitos de consumo de una población de dos millones de personas. Se pasa una encuesta a mil personas y se les pregunta si en su domicilio se cocina con gas, de los que 600 responden afirmativamente. Qué puedes afirmar sobre el número de personas en las que en su domicilio se usa gas con un nivel de confianza del 95 %.

Vídeo de un problema de selectividad resuelto: El tiempo de vida de los individuos de cierta especie animal tiene una distribución normal de media 8.8 meses y una desviación típica de 3 meses.



$a)$ ¿Qué porcentaje de individuos de esta espacie supera los 10 meses? ¿Qué porcentaje de individuos ha vivido entre 7 y 10 meses?

$b)$ Si se toman al azar 4 especímenes, ¿cuál es la probabilidad de que el menos uno no supere los 10 meses de vida?

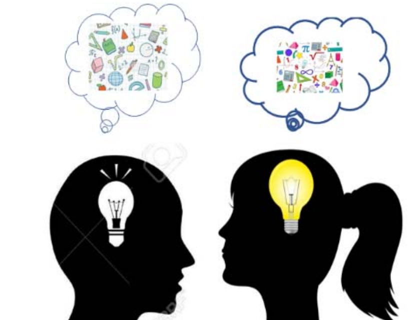
$c)$ ¿Qué valor de c es tal que el intervalo $(8.8 - c, 8.8 + c)$ incluye el tiempo de vida (medido en meses) del 98 % de los individuos de esta especie?

https://www.youtube.com/watch?v=E08i7gHUeaw

3. CONTRASTE DE HIPÓTESIS

3.1. Test de hipótesis. Contraste de hipótesis para la proporción poblacional

Empecemos con un ejemplo.



Actividad resuelta

- *La probabilidad de curarse una enfermedad con un cierto medicamento es* 0.68*. Se investiga un nuevo medicamento que queremos mejore el número de curaciones. Se tratan* 200 *enfermos de los que se curan* 150*. ¿Podemos estar seguros de que el nuevo medicamento es mejor que el antiguo?*

En primer lugar vamos a calcular la probabilidad de que con el primer medicamento se hubieran curado 150 enfermos. Tenemos una distribución binomial de media μ = np = 200·0.68 = 136, y $\sigma = \sqrt{npq} = \sqrt{200 \cdot 0.68 \cdot 0.32} = \sqrt{43.52} = 6.6$

Ajustamos la binomial con una normal, tipificamos y buscamos en la tabla:

$$P(z \geq \frac{150.5 - 136}{6.6} = 2.2) = 1 - P(z < 2.2) = 1 - 0.9861 = 0.0139$$

La probabilidad ha salido muy pequeña. Rechazamos la hipótesis. Aunque es posible que sí hubiera con el primer medicamento 150 curaciones, pero sólo en el 1.39 % de los casos.

Nivel de significación

En el ejemplo hemos partido de considerar cierta una hipótesis, que los medicamentos fueran iguales. Si la probabilidad sale menor que un cierto valor, llamado ***nivel de significación***, rechazamos la hipótesis.

Se suelen tomar como niveles de significación 5 %, 1 %, 0.1 % … según la naturaleza del problema.

En la actividad anterior diríamos que rechazamos la hipótesis de que ambos medicamentos sean igual de efectivos con un nivel de significación del 5 %, pero no podríamos rechazarla con un nivel de significación del 1 % por ser 1.39 > 1.

Actividad resuelta

- *En la actividad anterior, el número de curaciones observadas,* 150*, ¿es compatible con que el medicamento sea efectivo en el* 69 % *de los casos, con un nivel de significación del* 5 %*? ¿Y en el* 70 % *con igual nivel de significación?*

Repetimos el proceso para estos nuevos valores:

p	μ	σ	Z	$P(x \geq 150)$
0.69	np = 200·0.69 = 138	$\sigma = \sqrt{npq} = \sqrt{200 \cdot 0.69 \cdot 0.31}$ =6.49	$\frac{150.5-138}{6.5}$=1.92	1 - 0.9726 = 0.0274
0.7	np = 200·0.7 = 140	$\sigma = \sqrt{npq} = \sqrt{200 \cdot 0.7 \cdot 0.3}$ =6.48	$\frac{150.5-140}{6.48}$=1.62	1 - 0.9474 = 0.0526

Rechazamos la hipótesis con un nivel de significación del 5 % de que el porcentaje de curaciones sea del 68 %, y del 69 %. Esperamos que el porcentaje de curaciones del segundo medicamento sea superior al 70 %.

Test unilateral y test bilateral

Hemos considerado en las actividades anteriores la hipótesis de que ambos medicamentos tienen un porcentaje de curaciones iguales, y la hipótesis contraria de que el segundo medicamento tiene un mayor porcentaje de curaciones. Hemos calculado $P(x \geq 150)$, es decir, la probabilidad de que la variable aleatoria tome valores a la derecha de 150. Este tipo de test se denomina *unilateral*. Si debemos calcular probabilidades simétricas a ambos lados, se denomina *bilateral*.

Actividad resuelta

- *Queremos comprobar si una moneda no está trucada, con un nivel de significación del* 5 %. *Lanzamos la moneda al aire* 100 *veces y obtenemos* 60 *caras. ¿Aceptamos la hipótesis de que la moneda no está trucada?*

Tenemos las siguientes hipótesis:

H_0 = la moneda tiene una probabilidad de salir cara de 1/2.
H_1: La moneda está trucada, la probabilidad de cara es distinta de 1/2.
Es una distribución binomial de media $\mu = 100 \cdot (1/2) = 50$, y varianza $\sigma^2 = 100 \cdot (1/2) \cdot (1/2) = 25 \rightarrow \sigma = 5$.
La hipótesis H_1 indica que p podría ser mayor que 1/2 o menor que 1/2, por lo que debemos considerar tanto que se obtengan más de 50 caras como que se obtengan menos. Hemos obtenido 60 caras, que supera en 10 al valor medio, 50, luego vamos a calcular: $P(|x - 50| > 10)$, es decir,

$$P(x - 50 > 10) + P(x + 50 < -10)$$

De nuevo aproximamos la binomial con la normal:

$P(x > 60.5) + P(x < 39.5) = P(z > 2.1) + P(z < -2.1) = (1 - P(z \leq 2.1)) + (1 - P(z \leq 2.1)) = 2 \cdot (1 - P(z \leq 2.1)) = 2(1 - 0.9821) = 2(0.0179) = 0.0358$

Como 3.58 < 5, podemos rechazar la hipótesis de que la moneda esté equilibrada (tenga una probabilidad de 1/2, sea de *Laplace*:..) al nivel de significación del 5 %.

En ambos ejemplos tenemos duda sobre si el parámetro poblacional toma un valor determinado. Para salir de esa duda hacemos un test estadístico, tomando una muestra aleatoria que nos permita sacar conclusiones de la población, y aceptar o rechazar la hipótesis previamente emitida

Caso	H_0	H_1
Curaciones	$p = 69$ %	$p > 69$ %
Moneda	$p = 1/2$	$p \neq 1/2$

Actividades propuestas

25. Repite los cálculos de una actividad anterior para comprobar si una moneda no está trucada, con un nivel de significación del 5 %. Para ello lanzamos la moneda al aire 100 veces y obtenemos 65 caras. ¿Se puede asegurar que sea una moneda de probabilidad 1/2?

26. Se ha calculado que entre los deportistas que juegan al futbol hay un porcentaje de accidentes del 22 %. Se han estudiado el número de accidentes entre 400 personas que practican la natación y han resultados accidentadas 36 personas. ¿Es la natación igual de peligrosa que el futbol?



3.2. Contraste de hipótesis para la media poblacional

Podemos encontrar dos casos, que la hipótesis nula H_0 sea del tipo $\mu = \mu_0$, o que sea con una desigualdad: $\mu \geq \mu_0$, o bien $\mu \leq \mu_0$.

Paso 1: Hipótesis H_0: $\mu = \mu_0$, H_1: $\mu \neq \mu_0$.

Paso 2: Zona de aceptación.

Consideramos que la media se distribuye según $N(\mu_0, \frac{\sigma_0}{\sqrt{n}})$ lo que es cierto si la distribución poblacional es normal o si el tamaño de la muestra es suficientemente grande. La zona de aceptación de la hipótesis es el intervalo: $\mu \in \left(\mu_0 - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\quad \mu_0 + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$

Paso 3: Verificación: Se extrae la muestra y se calcula $\bar{x}$.

Paso 4: Decisión: Se acepta o se rechaza la hipótesis.

Actividad resuelta



- *Se piensa que el tiempo de renovación de un teléfono móvil, expresado en años, se puede aproximar mediante una distribución normal de media 2 y con desviación típica 0.4 años. Para contrastar esta hipótesis se pasa una encuesta a 100 personas, y el tiempo medio de renovación de sus teléfonos móviles ha sido de 1.8 años. ¿Se puede aceptar la hipótesis con un nivel de significación del 5 %?*

Paso 1: Hipótesis H_0: $\mu = \mu_0 = 2$, H_1: $\mu \neq 2$.

Paso 2: Zona de aceptación:

$$\left(\mu_0 - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\quad \mu_0 + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = \left(2 - 1.96 \cdot \frac{0.4}{\sqrt{100}},\quad 2 + 1.96 \cdot \frac{0.4}{\sqrt{100}}\right) = (1.92, 2.08)$$

Paso 3: Verificación:

Al extraer la muestra la media ha sido 1.8 que no pertenece al intervalo de aceptación.

Paso 4: Se rechaza la hipótesis de que la media sea 2.

Para el contraste unilateral la zona de aceptación no será simétrica.

Actividad resuelta

- *En la actividad anterior se quiere contrastar la hipótesis de que la media es superior a 2.*

Paso 1: Hipótesis H_0: $\mu = \mu_0 > 2$, H_1: $\mu \leq 2$.

Paso 2: Zona de aceptación: $\left(\mu_0 - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\quad +\infty\right) = \left(2 - 1.96 \cdot \frac{0.4}{\sqrt{100}},\quad \infty\right) = (1.92, +\infty)$.

Paso 3: Verificación: Al extraer la muestra la media ha sido 1.8 que no pertenece al intervalo de aceptación.

Paso 4: Se rechaza la hipótesis de que la media sea superior a 2.

3.3. Hipótesis nula. Error de primera y segunda especie

Hemos visto ejemplos en los que hemos hecho una hipótesis, (que ambos medicamentos eran iguales, que la moneda no estaba trucada…).

La hipótesis, H_0, de que no hay cambios se llama ***hipótesis nula***.

La hacemos con la intención de rechazarla, y aceptar la hipótesis contraria, H_1, de que sí hay cambios (el segundo medicamento es mejor, la moneda está trucada…).

Para decidir si rechazamos H_0 hemos fijado un nivel de significación, α, hemos realizado un test que nos suministra una zona crítica D en la que:

Si suponiendo que H_0 es verdadera ocurre que $P(x \in D) < \alpha$, entonces rechazamos H_0.

Si suponiendo que H_0 es verdadera ocurre que $P(x \in D) \geq \alpha$, entonces NO rechazamos H_0.

Podemos cometer dos tipos de errores, el error de tipo 1, es rechazar H_0 siendo verdadera; y el error del tipo 2, de aceptar H_0 siendo falsa.

La probabilidad de cometer un error del primer tipo es el nivel de significación α.

Actividades propuestas

27. La tasa de natalidad de una región ha sido del 8.7 por mil habitantes durante un cierto año. Suponemos que la tasa de natalidad es la misma al año siguiente, ¿hasta qué número de nacimientos entre 3 000 habitantes estarías dispuesto a confirmar dicha hipótesis?

3.4. Analogía entre intervalos de confianza y contraste de hipótesis

Existe una gran relación entre el intervalo de confianza para un parámetro de una distribución y el contraste de hipótesis sobre el mismo. Si al construir el intervalo de confianza, el estimador muestral no pertenece a él, se rechaza la hipótesis nula de que la población tenga dicho parámetro.

El nivel de significación α de un contraste de hipótesis es el complementario del nivel de confianza de una estimación: $1 - \alpha$.

CURIOSIDADES. REVISTA

EL EFECTO PLACEBO Y EL EFECTO NOCEBO

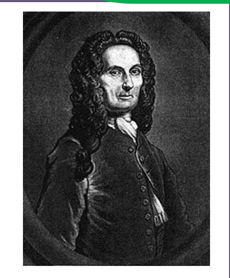
Antes de que un medicamento pueda comercializarse debe superar una serie de estrictas pruebas que arrojen seguridad acerca de su eficacia curativa.

Una de las pruebas más comunes consiste en seleccionar una muestra de enfermos y dividirlos aleatoriamente en dos grupos; un grupo recibe el medicamento, y el otro, sin saberlo, una sustancia en apariencia igual, pero sin ningún poder terapéutico: un placebo.

De esta forma, al final del ensayo pueden compararse los resultados entre los dos grupos y determinar la eficacia del medicamento. Para ello se emplean herramientas estadísticas como la correlación.

Sorprendentemente, hay un número significativo de pacientes que, habiendo recibido el placebo, mejoran de forma ostensible. Por ejemplo, esta contrastado que, en muchas enfermedades relacionadas con el dolor, entre el 10 % y el 15 % de los pacientes experimenta un alivio notable habiendo seguido un tratamiento exclusivamente de placebo.





Moivre



Laplace



Gauss

Distribución Normal

La **importancia** de esta distribución se debe a que se utiliza para modelar numerosos fenómenos naturales, médicos y sociales. Son fenómenos en los que influyen muchas variables difíciles de controlar, por lo que podemos suponer que es suma de distintas causas independientes.

Ejemplos clásicos de fenómenos que se distribuyen según una normal son:

- Fenómenos morfológicos como la estatura o el peso
- Fisiológicos como los efectos de un fármaco
- Sociológicos como los de consumo
- Psicológicos como el cociente intelectual
- El ruido en las telecomunicaciones
- Los errores cometidos al medir una magnitud…

La **historia** de la distribución normal. Aparece por primera vez con *Abraham de Moivre* en un artículo publicado en 1733, sobre la distribución binomial para valores grandes de n.

El resultado fue trabajado por *Laplace* en su libro sobre la Teoría de las probabilidades trabajando sobre errores.

También sobre errores la utilizó *Gauss*, analizando datos astronómicos. En su honor también se denomina a la curva normal, *campana de Gauss*.

RESUMEN

Muestra aleatoria simple	Todos los individuos de la población tienen la misma probabilidad de ser elegidos en la muestra.	Se numera la población y se usan números aleatorios para elegir la muestra.
Teorema central del límite	Si X es una variable aleatoria de una población de media μ finita y desviación típica σ finita. Entonces: La distribución de la media muestral de tamaño n tiene de media μ y desviación típica $\frac{\sigma}{\sqrt{n}}$ y se aproxima a una distribución normal a medida que crece el tamaño de la muestra	Población $N(10, 2)$ Muestra de tamaño $n = 100$. $\rightarrow$ Distribución de la media muestral: $N(10, 0.2)$
Media muestral	$N(\bar{x}, \frac{\sigma}{\sqrt{n}})$	$P(8 < \bar{x} < 12) = P(-1 < z < 1) = 2P(z < 1) - 1 = 0'6832$
Proporción muestral	$N(p, \sqrt{\frac{p(1-p)}{n}})$	Proporción: 5 %. Muestra de tamaño $n = 100 \rightarrow N(0.05, 0.03)$
Intervalo de confianza	**Intervalo de confianza**: Si $P(a < X < b) = 0'95$ tenemos el intervalo de confianza (a, b) **Nivel de confianza o coeficiente de confianza:** $1 - \alpha = \gamma$, en nuestro ejemplo, 0.95 **Nivel de significación o de riesgo:** α, en nuestro ejemplo, 0.05 **Valor crítico:** k_1 y k_2, que dejan a la derecha (o a la izquierda) un área $\alpha/2$. En la $N(0, 1)$ son -1.96 y 1.96 para $\alpha = 0.05$. **Margen de error:** Diferencia entre los extremos del intervalo de confianza.	
Intervalo de confianza para la media	$\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$	$N(2, 1)$, $1 - \alpha = \gamma = 0.95$; $P(-1.96 < (X-2)/1 < 1.96) = 0.95$ $\Rightarrow P(1.8 < X < 2.2) = 0.95$
Error máximo admisible. Tamaño mínimo de la muestra	$E = z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow n = \left(z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{E}\right)^2$	$N(2, 1)$, $1 - \alpha = 0.95$; $n = 100$ $E = 1.96 \cdot (1/10) = 0.196$. Si $E = 0.5 \rightarrow n = (1.96 \cdot (1/0.5))^2 \approx 16$
Intervalo de confianza para la proporción	$p \in \left(\hat{p} - z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}},\ \hat{p} + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p} \cdot (1-\hat{p})}{n}}\right)_{\gamma}$	Proporción:1/6. Muestra de tamaño $n = 120$. $1 - \alpha = 0.95 \rightarrow z_{1-\alpha/2} = 1.645$; s = 0.034 $\rightarrow (0.111, 0.223)$
Contraste de hipótesis	Paso 1: Hipótesis H_0: $\mu = \mu_0$, H_1: $\mu \neq \mu_0$. Paso 2: Zona de aceptación. $\mu \in \left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$ Paso 3: Verificación: Se extrae la muestra y se calcula $\bar{x}$. Paso 4: Decisión: Se acepta o se rechaza la hipótesis.	



EJERCICIOS Y PROBLEMAS

1. Utiliza las tablas de la normal estándar y comprueba las probabilidades siguientes:

a) $P(z < 1) = 0.8413$; b) $P(z \leq 0.7) = 0.7580$; c) $P(z > 1) = 1 - 0.8413 = 0.1587$; d) $P(z \geq 1.86) = 0.0314$;

e) $P(-1.83 < z < -1) = 0.1251$; f) $P(z > 1.38) = 0.0838$; g) $P(-1.83 \leq z < 0.75) = 0.7398$.

2. Utiliza las tablas de la normal estándar para calcular las probabilidades siguientes:

a) $P(z < 0.72)$; b) $P(z \leq 1.21)$; c) $P(z > 0.93)$; d) $P(z \geq -1.86)$;

e) $P(-1{,}02 < z < -0.85)$; f) $P(0.65 < z < 1.42)$; g) $P(1.76 > z > 0.72)$; h) $P(-0.9 > z > -0.51)$.

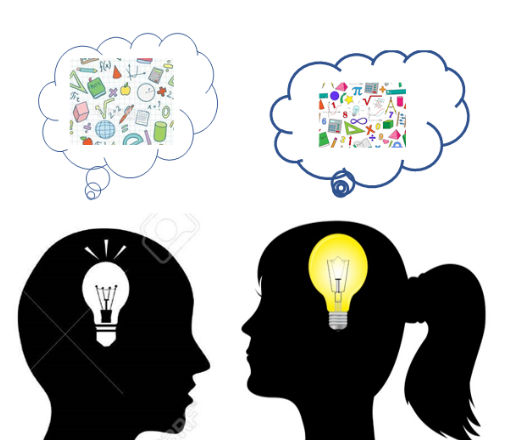
3. Una variable aleatoria X sigue una distribución normal de media 5 y desviación típica 0.5. Calcula las siguientes probabilidades:

a) $P(X < 6)$; b) $P(X \leq 4)$; c) $P(X > 3)$; d) $P(X \geq 5.5)$;

e) $P(-3 < X < -1)$; f) $P(X > 2)$; g) $P(3 \leq X < 7)$; h) $P(6 > X > 2)$.

4. En un centro escolar hay 900 estudiantes, que son 600 de ESO y 300 de Bachillerato. Se quiere tomar una muestra aleatoria por muestreo estratificado proporcional de tamaño 50. ¿Cuántos estudiantes se deben escoger de forma aleatoria de ESO y cuántos de bachillerato?

5. El número de megabytes (Mb) descargados mensualmente por un grupo de clientes de una compañía de telefonía móvil se aproxima por una distribución normal con media 4 Mb y desviación típica igual a 1.5 Mb. Se toma una muestra aleatoria simple de tamaño 64.



a) ¿Cuál es la probabilidad de que la media muestra sea inferior a 3.5 Mb?

b) ¿Sea superior a 4.5 Mb?

c) Se supone ahora que la media poblacional es desconocida y que la media muestra toma el valor 3.7 Mb. Obtén un intervalo de confianza al 95 % para la media de la población. Obtén también un intervalo de confianza al 99 % para la media de la población. ¿Es mayor o menos que el anterior? Explica este resultado

6. La duración en horas de un cierto tipo de bombillas de bajo consumo se puede aproximar por una distribución normal de media μ y desviación típica igual a 3600 horas. Se toma una muestra aleatoria simple.

a) ¿Qué tamaño muestral se necesitaría como mínimo para que, con un nivel de confianza del 95 %, el valor absoluto de la diferencia entre μ y la duración media observada $\overline{X}$ de esas bombillas sea inferior a 100 horas?

b) Si el tamaño de la muestra es 121 y la duración media observada $\overline{X}$ es de 4000 horas, obtén un intervalo de confianza al 95 % para la media poblacional μ.

7. La longitud, en milímetros (mm), de los individuos de una determinada plantación de mejillones se puede aproximar por una variable aleatoria con distribución normal de media desconocida μ y desviación típica igual a 3 mm.

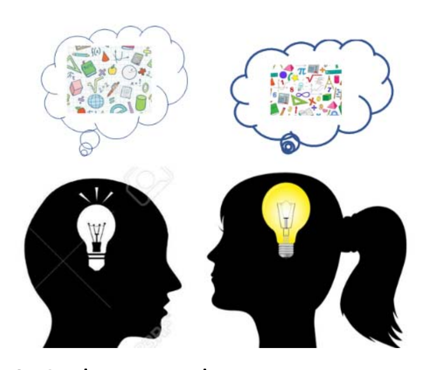
a) Se toma una muestra aleatoria simple de 64 mejillones y se obtiene una media muestral igual a 70 mm. Determina un intervalo de confianza para la media poblacional de la longitud de los mejillones con un nivel de confianza del 99 %. Determina también un intervalo de confianza para la media poblacional de la longitud de los mejillones con un nivel de confianza del 95 %.

b) Determina el tamaño muestral mínimo necesario para que el error máximo cometido en la estimación de μ por la media muestral sea menor o igual que 5 mm con un nivel de confianza del 95 %.

8. El consumo mensual de leche (en litros) de los alumnos de un determinado colegio se puede aproximar por una variable aleatoria con distribución normal de media μ y desviación típica σ = 3 litros.

a) Se toma una muestra aleatoria simple y se obtiene el intervalo de confianza (16; 20) para estimar μ, con un nivel de confianza del 95 %. Calcula la media muestral y el tamaño de la muestra elegida.

b) Se toma una muestra aleatoria simple de tamaño 81. Calcula el error máximo cometido en la estimación de μ mediante la media muestral con un nivel de confianza del 95 %.



9. El consumo familiar diario de electricidad (en kW) en cierta ciudad se puede aproximar por una variable aleatoria con distribución normal de media μ = 6.3 kW y desviación típica 0.9 kW. Se toma una muestra aleatoria simple de tamaño 100. Calcula:

a) La probabilidad de que la media muestral esté comprendida entre 6 kW y 6.6 kW.

b) El nivel de confianza con el que se ha calculado el intervalo de confianza (6.1; 6.6) para la media del consumo familiar diario.

10. Se ha tomado una muestra aleatoria simple de 9 pacientes y se ha anotado el número de días que han recibido tratamiento para trastornos digestivos que sufren. Los resultados han sido:

100, 98, 75, 103, 84, 95, 105, 82, 107.

Se sabe que la duración, en días, del tratamiento se puede aproximar por una variable aleatoria con distribución normal de media μ desconocida y desviación típica 9 días.

a) Determina un intervalo de confianza con un nivel del 95 % para μ.

b) ¿Qué tamaño mínimo debe tener la muestra para que el error máximo cometido en la estimación de la media sea menor de 5 días, con un nivel de confianza del 95 %?

11. El tiempo de renovación de un teléfono móvil, expresado en años, se puede aproximar mediante una distribución normal con desviación típica 0.2 años.

a) Se toma una muestra aleatoria simple de 81 usuarios y se obtiene una media muestral igual a 1.8 años. Determina un intervalo de confianza al 95 % para el tiempo medio de renovación de un teléfono móvil.

b) Determina el tamaño muestral mínimo necesario para que el valor absoluto de la diferencia entre la media muestral y la media poblacional sea menor o igual a 0.03 años con un nivel de confianza del 95 %.

12. Se considera una variable aleatoria con distribución normal de media μ y desviación típica igual a 1.2. Se toma una muestra aleatoria simple de 100 elementos.

a) Calcula la probabilidad de que el valor absoluto de la diferencia entre la media muestral y μ sea mayor o igual que 4.

b) Determina un intervalo de confianza del 90 % para μ; si la media muestral es igual a 50.

13. La estatura en centímetros (cm) de los varones mayores de edad de una determinada población se puede aproximar por una variable aleatoria con distribución normal de media μ y desviación típica σ = 15 cm.

a) Se toma una muestra aleatoria simple de 100 individuos obteniéndose una media muestral $\bar{x}$ = 174 cm. Determina un intervalo de confianza al 95 % para μ.

b) ¿Cuál es el mínimo tamaño muestral necesario para que el error máximo cometido en la estimación de μ por la media muestral sea menor que 5 cm, con un nivel de confianza del 90 %?

14. El mínimo tamaño muestral necesario para estimar la media de una determinada característica de una población que puede aproximarse por una variable aleatoria con distribución normal de desviación típica σ, con un error máximo de 2.27 y un nivel de confianza del 90 %, supera en 1000 unidades al que se necesitaría si el nivel de confianza fuera del 95 % y el error máximo fuera de 5.23. Expresa los tamaños muestrales en función de la desviación típica σ y calcula la desviación típica de la población y los tamaños muestrales respectivos.

AUTOEVALUACIÓN

1. Indica cuál de los siguientes motivos no es por el que se recurre a una muestra:

a) El proceso de medición es destructivo

b) La población es muy numerosa

c) La población es imposible o difícil de controlar

d) La población tiene mal carácter

2. Una ganadería tiene diez mil ovejas de diferentes razas. Queremos extraer una muestra de 100 ovejas. Indica el tipo de muestreo más adecuado:

a) muestreo aleatorio sistemático b) muestreo aleatorio estratificado

c) muestreo no aleatorio d) muestreo aleatorio por conglomerados

3. Indica cuál de las siguientes afirmaciones es falsa en una distribución $N(0, 1)$:

a) $P(z < 0) = 1$ b) $P(z < 0) = 0.5$ c) $P(z = \sigma) = 0$ d) $P(z > 0) = 0.5$.

4. De una población de media 69 y desviación típica 8 se toma una muestra de tamaño 12. La probabilidad de que un individuo de la muestra tenga un valor mayor que 93 es:

a) $P(x > 93) = 0.9987$ b) $P(x > 93) = 0.6501$ c) $P(x > 93) = 0.1293$ d) $P(x > 93) = 0.0013$.

5. Los parámetros de una distribución son $\mu = 10$ y desviación típica $\sigma = 20$. Se extrae una muestra de 100 individuos. El valor de $P(8 < \bar{x} < 12)$ es:

a) $P(z < 1) = 0.8416$ b) 0.6838 c) 0.3168 d) 0.1584.

6. En el control de calidad de una fábrica de chocolate, se envasan tabletas de 100 gramos con una desviación típica de 2 gramos. Se toma una muestra de 50 tabletas. Calcula la probabilidad de que el peso medio de las tabletas sea menor que 99 gramos:

a) 0.0002 b) 0.9998 c) 0.3541 d) 0.0023.

7. En el control de calidad de una envasadora de estuches de jamón, se envasan en estuches de 100 gramos con una desviación típica de 2 gramos. La probabilidad de que un lote de 400 estuches pese más de 40100 gramos es de:

a) 0.9938 b) 0.0062 c) 0.0002 d) 0,9998

8. Determina un intervalo de confianza con un nivel de confianza del 0.95 de una $N(2, 0.1)$:

a) $P(1.8 < X < 2.2) = 0.95$ b) $P(1.9 < X < 2.1) = 0.95$ c) $P(1.8 < X < 2.2) = 0.99$ d) $P(1 < X < 2) = 0.90$

9. Se ha elegido una muestra aleatoria simple de 1000 componentes y en ella se ha obtenido que la proporción de defectuosos es del 3.7 %. Determina el intervalo de confianza al 99 % para la proporción de componentes defectuosos que se producen en una fábrica:

a) (0.0371, 0.0375) b) (0.0258, 0.0351) c) (0.0216, 0.0524) d) (0.0111, 0.0222)

10. ¿Cuál debe ser el tamaño de la muestra en una población de 8 millones de votantes para conocer si tienen la intención de votar a un determinado partido político con una probabilidad de acierto del 0.95 y un margen de error inferior a 0.02?:

a) 2401 b) 1959 c) 2502 d) 3026

ÁREAS BAJO LA DISTRIBUCIÓN DE PROBABILIDAD NORMAL ESTÁNDAR, N(0, 1)

Y
$P(z \leq k)$
N(0, 1)
X
k

Tabla de la uam: Universidad Autónoma de Madrid

z	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
4,0	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000



PROBLEMAS PROPUESTOS EN SELECTIVIDAD

Puedes ver muchos problemas de selectividad, de distintos años y diferentes comunidades autónomas, resueltos en la web de Marea Verde, en SELECTIVIDAD.

También en "Problemas resueltos por el alumnado" tienes problemas de estimación resueltos por estudiantes de un instituto. Algunos problemas resueltos por el alumnado.

1. El número de megabytes (Mb) descargados mensualmente por un grupo de clientes de una compañía de telefonía móvil con la tarifa AA se puede aproximar por una distribución normal con media 3.5 Mb y desviación típica igual a 1.5 Mb. Se toma una muestra aleatoria simple de tamaño 49. a) ¿Cuál es la probabilidad de que la media muestral sea inferior a 3.37 Mb? b) Supóngase ahora que la media poblacional es desconocida y que la media muestra toma el valor 3.42 Mb. Obténgase un intervalo de confianza al 95 % para la media de la población.

2. La duración en horas de un cierto tipo de bombillas se puede aproximar por una distribución normal de media μ y desviación típica igual a 1940 horas. Se toma una muestra aleatoria simple.

a) ¿Qué tamaño muestral se necesitaría como mínimo para que, con un nivel de confianza del 95 %, el valor absoluto de la diferencia entre μ y la duración media observada $\overline{X}$ de esas bombillas sea inferior a 100 horas?

b) Si el tamaño de la muestra es 225 y la duración media observada $\overline{X}$ es de 12415 horas, obténgase un intervalo de confianza al 90 % para μ.

3. La longitud, en milímetros (mm), de los individuos de una determinada colonia de gusanos de seda se puede aproximar por una variable aleatoria con distribución normal de media desconocida μ y desviación típica igual a 3 mm.

a) Se toma una muestra aleatoria simple de 48 gusanos de seda y se obtiene una media muestral igual a 36 mm. Determínese un intervalo de confianza para la media poblacional de la longitud de los gusanos de seda con un nivel de confianza del 95 %.

b) Determínese el tamaño muestral mínimo necesario para que el error máximo cometido en la estimación de μ por la media muestral sea menor o igual que 1 mm con un nivel de confianza del 90 %.

4. El consumo mensual de leche (en litros) de los alumnos de un determinado colegio se puede aproximar por una variable aleatoria con distribución normal de media μ y desviación típica $\sigma = 3$ litros.

a) Se toma una muestra aleatoria simple y se obtiene el intervalo de confianza (16.33; 19.27) para estimar μ, con un nivel de confianza del 95 %. Calcúlese la media muestral y el tamaño de la muestra elegida.

b) Se toma una muestra aleatoria simple de tamaño 64. Calcúlese el error máximo cometido en la estimación de μ mediante la media muestral con un nivel de confianza del 95 %.

5. El consumo familiar diario de electricidad (en kW) en cierta ciudad se puede aproximar por una variable aleatoria con distribución normal de media μ y desviación típica 1.2 kW. Se toma una muestra aleatoria simple de tamaño 50. Calcúlese:
a) La probabilidad de que la media muestral esté comprendida entre 6 kW y 6.6 kW, si $\mu = 6.3$ kW.
b) El nivel de confianza con el que se ha calculado el intervalo de confianza (6.1; 6.9) para la media del consumo familiar diario.

6. Se ha tomado una muestra aleatoria simple de diez pacientes y se ha anotado el número de días que han recibido tratamiento para los trastornos del sueño que sufren. Los resultados han sido:

290; 275; 290; 325; 285; 365; 375; 310; 290; 300.

Se sabe que la duración, en días, del tratamiento se puede aproximar por una variable aleatoria con distribución normal de media μ desconocida y desviación típica 34.5 días.

a) Determínese un intervalo de confianza con un nivel del 95 % para μ.

b) ¿Qué tamaño mínimo debe tener la muestra para que el error máximo cometido en la estimación de la media sea menor de 10 días, con un nivel de confianza del 95 %?

7. El tiempo de renovación de un teléfono móvil, expresado en años, se puede aproximar mediante una distribución normal con desviación típica 0.4 años.

a) Se toma una muestra aleatoria simple de 400 usuarios y se obtiene una media muestral igual a 1.75 años. Determínese un intervalo de confianza al 95 % para el tiempo medio de renovación de un teléfono móvil.

b) Determínese el tamaño muestral mínimo necesario para que el valor absoluto de la diferencia entre la media muestral y la media poblacional sea menor o igual a 0.02 años con un nivel de confianza del 90 %.

8. Se considera una variable aleatoria con distribución normal de media μ y desviación típica igual a 210. Se toma una muestra aleatoria simple de 64 elementos.

a) Calcúlese la probabilidad de que el valor absoluto de la diferencia entre la media muestral y μ sea mayor o igual que 22.

b) Determínese un intervalo de confianza del 99 % para μ; si la media muestral es igual a 1532.

9. La estatura en centímetros (cm) de los varones mayores de edad de una determinada población se puede aproximar por una variable aleatoria con distribución normal de media μ y desviación típica σ = 16 cm.

a) Se tomó una muestra aleatoria simple de 625 individuos obteniéndose una media muestral $\bar{x}$ = 169 cm. Hállese un intervalo de confianza al 98 % para μ.

b) ¿Cuál es el mínimo tamaño muestral necesario para que el error máximo cometido en la estimación de μ por la media muestral sea menor que 4 cm, con un nivel de confianza del 90 %?

10. El mínimo tamaño muestral necesario para estimar la media de una determinada característica de una población que puede aproximarse por una variable aleatoria con distribución normal de desviación típica σ, con un error máximo de 3.290 y un nivel de confianza del 90 %, supera en 7500 unidades al que se necesitaría si el nivel de confianza fuera del 95 % y el error máximo fuera de 7.840: Exprésense los tamaños muestrales en función de la desviación típica σ y calcúlense la desviación típica de la población y los tamaños muestrales respectivos.

Nota: Utilícese $z_{0;05}$ = 1.645.